LLMs effectively capture cultural, technical, and contextual nuances for rabies education in Sepedi.

Exploring Zoonotic Disease Knowledge Through AI for Enhanced Risk, Prevention, and Response Awareness in Low-Resource Languages

Reolin Govender, Armand de Wet

INTRO

Rabies awareness remains limited in affected communities due to inadequate language inclusivity. This project targets Sepedi, a low-resource language, to evaluate how well Large Language Moßßdels (LLMs) communicate culturally and contextually relevant rabies information.

METHODS

Dataset Creation → Model Fine-tuning → Testing & Validation → Visualisation & Deployment

- **Dataset Creation:**
  Created 60 rabies Q&A pairs in English, translated to Sepedi using Gemini, and refined by a Sepedi expert.
- **Model Fine-tuning:**
  Fine-tuned GPT-4o and Gemini-1.5 models using the refined dataset.
- **Testing and Validation:**
  Generated additional Sepedi questions; evaluated model outputs with expert human validation for fluency, accuracy, and contextual fit.
- **Visualisation & Deployment:**
  Built an interactive chatbot with Streamlit (Figure 2) to compare model outputs and demonstrate real-time engagement.

RESULTS

- Both fine-tuned models—GPT-4o-2024-08-06 and Gemini-1.5-Flash-001—were able to generate Sepedi responses that were **fluent, factually accurate, and culturally appropriate**, as confirmed by external human validation (using criteria in Table 1).
- Fine-tuned models outperformed translation-based approaches in clarity and relevance.
- Responses remained consistent across paraphrased and unseen prompts.

OpenAI ⚡ Gemini

VISUALISATION

- Developed using Streamlit for fast prototyping and public deployment.
- Users can chat with the models in Sepedi and receive real-time answers (Figure 2).
- Includes side-by-side comparison of GPT-4o and Gemini outputs.
- Displays response time metrics to evaluate speed and usability (Figure 3).
- Word cloud summarises model vocabulary and thematic focus (Figure 4).

DISCUSSION

- LLMs show promise for inclusive, culturally aware health education.
- Careful dataset preparation is key for low-resource languages.
- Future work includes community validation, data expansion, and real-world deployment.

NEXT STEPS

- Expand and diversify training data.
- Involve communities for broader validation.
- Add features like voice input, offline access, and feedback.

This pilot supports the use of AI to close language gaps in public health education.

Criterion	1 (Poor)	2 (Fair)	3 (Good)	4 (Very Good)	5 (Excellent)
Fluency / Grammar	Unnatural, many errors	Frequent grammar or phrasing issues	Understandable but some awkward parts	Smooth with minor issues	Native-like, fluent and natural
Factual Accuracy	Incorrect or misleading	Several factual errors	Mostly correct but minor inaccuracies	Almost entirely accurate	Completely accurate
Relevance to Prompt	Off-topic or unrelated	Partially addresses the question	Mostly relevant	Very relevant	Directly and completely answers the prompt
Clarity / Understandability	Confusing or incoherent	Difficult to follow	Understandable with effort	Clear and mostly easy to follow	Very clear and easily understood
Cultural & Contextual Appropriateness	Offensive or inappropriate	Culturally out of touch or awkward	Mostly appropriate	Suitable and respectful	Culturally aligned, respectful, and sensitive

Table 1: Human Validation Criteria

Human Validation Results

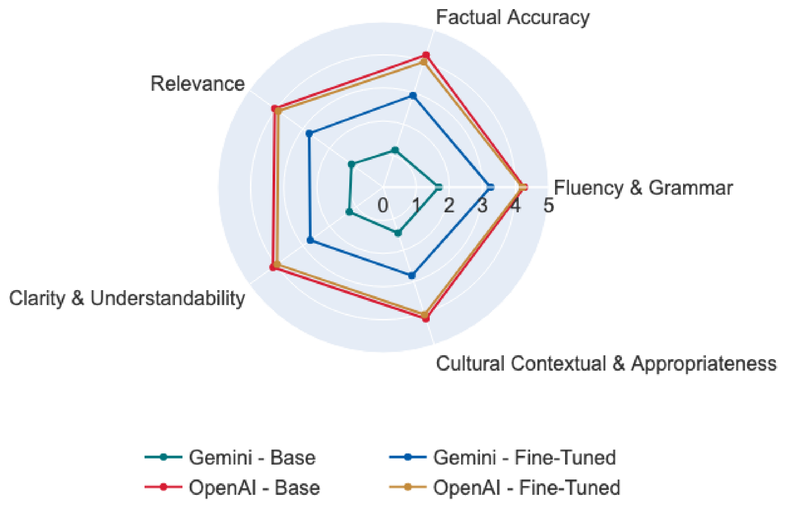

Figure 1: Radar Plot of Human Validation Scores

Visualisation and Deployment

Figure 2: Chatbot Interface: Side-by-Side LLM Comparison

Figure 3: Response Time Comparison of GPT-4o and Gemini Models

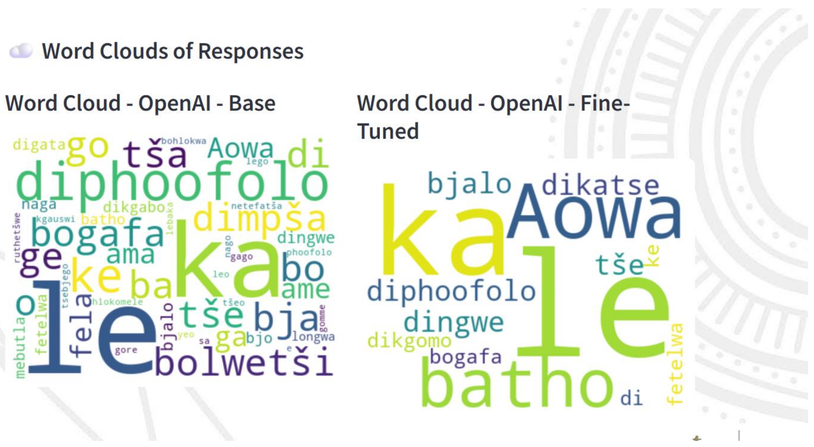

Figure 4: WordClouds of LLMs Responses

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Department of Computer Science

Faculty of Engineering, Built Environment and Information Technology
Fakulteit Ingenieurswese, Bou-omgewing en Inligtingtegnologie / Lefapha la Boetšenere, Tikologo ya Kago le Theknolotši ya Tshedimošo

Capstone Project - MIT 808

Course Coordinators:
Dr. Vukosi Marivate (vukosi.marivate@cs.up.ac.za)
Abiodun Modupe (abiodun.modupe@cs.up.ac.za)

Scan me